Locke, Nozick and the state of nature

Bruner, Justin P.

Published in:
Philosophical Studies

DOI:
10.1007/s11098-018-1201-9

Publication date:
2020

Philos Stud (2020) 177:705–726
https://doi.org/10.1007/s11098-018-1201-9

Locke, Nozick and the state of nature

Justin P. Bruner[1]

Published online: 20 November 2018

Abstract Recently, philosophers have drawn on tools from game theory to explore behavior in Hobbes' state of nature (Vanderschraaf in Econ Philos 22:243–279, 2006; Chung in J Am Philos Assoc 1:485–508, 2015). I take a similar approach and argue the Lockean state of nature is best conceived of as a conflictual coordination game. I also discuss Nozick's famous claim regarding the emergence of the state and argue the path to the minimal state is blocked by a hitherto unnoticed free-rider problem. Finally, I argue that on my representation of the Lockean state of nature both widespread conflict and lasting peace are possible. This, I contend, is in line with one popular interpretation of Locke (Simmons in Polit Theory 17:449–470, 1989).

Keywords Social contract · Game theory · State of nature · John Locke · Robert Nozick

1 Introduction

What is the outcome of life in the state of nature (SON)? Social contract theorists have long puzzled over this question, although they have provided vastly different answers. Hobbes assumes individuals are, as Gregory Kavka puts it, predominantly egoistic, and life is 'brutish and short,' to say the least. This characterization stands in stark contrast with the Lockean SON. Locke assumes a moralized SON in which tranquility and peace are possible and individuals are obliged to obey the law of nature.

✉ Justin P. Bruner
j.p.bruner@rug.nl

[1] Department of Theoretical Philosophy and Centre for Philosophy, Politics and Economics, University of Groningen, Groningen, The Netherlands

Recently, a variety of authors have imported tools from the social sciences to understand social contract formation and behavior in the SON. Peter Vanderschraaf, for instance, has argued with the support of a game-theoretic model that behavior in the Hobbesian SON will inevitably dissolve into a war of all against all (Vanderschraaf 2006). This is reinforced by more recent work by Hun Chung (2015). Despite the healthy interest in Hobbes' account, tools from game theory have yet to be used to represent and understand the dynamics of Locke's SON.[1] Since Locke and Hobbes have different characterizations of the SON, a careful game-theoretic exploration is called for in order to determine in what way, if at all, the outcome of the Lockean and Hobbesian SON differ. We argue that the Lockean SON is best represented as a conflictual coordination game involving disagreement among morally motivated individuals over how to best interpret and apply the law of nature. Disagreement of this kind can result in violent conflict and is one of the primary 'inconveniences' of the SON. We show that on our representation of the Lockean SON, conflict can be by-and-large avoided so long as individuals are not inclined to disregard the law of nature. This is in line with one interpretation of Locke due to John Simmons (1989), who argues Locke's seemingly contradictory characterizations of the SON as a place of peace as well as a place of enmity correspond to scenarios in which a minority or a majority of individuals disregard the law of nature, respectively.

Finally, a better understanding of Locke sheds light on the writings of Robert Nozick, who famously put forth an account of how individuals in the Lockean SON can inadvertently 'back into a state.' Some commentators have argued that the account, if successful, provides a powerful justification of the libertarian minimal state (Gaus 2011a, b), as it would demonstrate how a state need not arise via egregious rights violations, but instead through an invisible-hand process involving the decisions of autonomous individuals interested in securing their rights and safety.[2] We show how the progression to the minimal state is stalled as the result of an hitherto unnoticed free-rider problem.

This paper proceeds as follows. We begin with Locke's description of the SON and then discuss Nozick's origin story. In Sect. 3 we discuss how to best represent the SON game-theoretically before arguing the transition from the SON to minimal state is thwarted by a free-rider problem. In Sect. 4, we argue that on our representation of the SON, both peace and widespread war are possibilities.

[1] Only Vanderschraaf (2006) and Kavka (manuscript) have modeled the Lockean SON using tools from game theory. Neither provide a detailed discussion or development of the Lockean SON. More recently, Kogelmann and Ogden (2018) and Chung (manuscript) have provided formalizations of the so-called Lockean proviso.

[2] For Nozick, the minimal state is a state which restricts its activities to the protection of rights, property and contract.

2 From Locke to the minimal state

According to Locke, individuals have rights which precede any social contract as well as a set of duties emanating from the immutable 'law of nature'. Individuals have a right to life, liberty and property. Such 'natural rights' are granted at birth and all have such rights equally. Locke further contends the use of reason leads individuals to recognize the law of nature (LN), which stipulates individuals have certain duties to refrain from harming others. Locke posits that the LN plays a role in governing behavior, noting the SON is a 'state of perfect freedom to order actions and dispose of their possessions and persons as they think fit within the bounds of the law of nature' (Locke 1690/1980: 4). Additionally, individuals who have suffered harm at the hands of another have the right to exact compensation from transgressors. This 'executive power' is rather broad, as Locke thinks all have the right to punish those who violate the LN in order to deter future rights violations and harm.

Individuals find themselves in the Lockean SON when they lack a 'common superior on earth, to judge between them' (Locke 1690/1980: 19). Individuals must navigate disputes they have with others without an appeal to a 'common judge with authority' (Locke 1690/1980: 19). Locke's definition of the SON is rather minimal, and thus consistent with a variety of social characterizations. For instance, the SON could be characterized as a somewhat primitive place involving no production or possessions. Locke's definition is also consistent with a more elaborate imagining of the SON, replete with 'property in land, money, commerce and cities' (Simmons 1989, p. 458). What these instantiations of the SON share is the presence of a particular problem. Namely, individuals in all versions of the SON *must stand as judge in their own case*. When one individual harms another, they must attempt to rectify the situation on their own. Locke takes this to be problematic, however, since individuals are biased by self-love. Thus, an individual cannot help but be partial when assessing her own case. This leads to disagreement between well-meaning individuals attempting to abide by the LN. For instance, victims may demand excessive punishment or compensation for the harm they have suffered while the perpetrator may downplay the damage. Such disagreements can lead to violent conflict and disorder. This strife is the major inconvenience of the SON and primary reason Locke thinks individuals will seek the protection of a state.

Nozick's account of the SON in *Anarchy, State and Utopia* (ASU) is very similar to the one given above. That said, there are minor differences worth emphasizing. For instance, Nozick takes as his starting point a particularly rosy version of the Lockean SON, and assumes a 'preponderant majority, though not all, of the persons living in the state of nature' act in accordance with Locke's LN (Nozick 1974, p. 17). Following Locke, Nozick nonetheless acknowledges that even in this idyllic setting problems may emerge. Disagreements between agents in the SON occur since the natural law must be interpreted and does not provide for every contingency (Nozick 1974, p. 11). Individuals will inevitably allow self-love to bias their judgments when exercising executive power. This results in the overestimation of the amount of harm or damage suffered. With no common judge with authority,

individuals will demand excessive compensation and the private enforcement of rights will likely lead to brutual feuds and an 'endless series of acts of retaliation and exactions of compensation' (Nozick 1974, 11). It is worth stressing just how costly Nozick thinks such feuds will likely be. Once a feud has begun, there will be little chance of settling and ending the dispute peacefully. The parties will be stuck in a long cycle of retaliations as both sides will feel mutually wronged. Hence, the disagreements which inevitably occur when men must act as judge in their own case can easily escalate to brutal and violent conflict. Similar to Locke, Nozick takes this to be the primary inconvenience of the SON.

Nozick departs from Locke, however, when considering how this inconvenience could be dealt with. While Locke assumes the best course of action is to flee the SON for the protection and comfort of civil society, Nozick considers arrangements within the SON that could potentially ensure peace and stability. This motivates Nozick's discussion of protection associations and protection agencies, to which we now turn.

Protection associations are loosely connected groups of individuals who come to the aid of fellow group-members. An individual supports another because either they have previously received help from the aggrieved individual or they wish to secure help from her in the future. Nozick is skeptical such associations can function successfully because conflict among members may destabilize the association. Instead, Nozick argues a division of labor will emerge, whereby a subset of people are hired to provide protective services and enforce the rights of others. These individuals attempt to sell these protective services to others in the SON. When the rights of their clients are violated, they are responsible for pursuing compensation and punishing the guilty party[3]. These protection agencies compete for clients, and, furthermore, can come into violent conflict with other agencies. Eventually, competition and conflict weed out inefficient protection agencies and (nearly) all individuals in a particular geographic location come to patronize a 'dominant protection agency.' This agency, referred to by Nozick as the ultra-minimal state, is a monopoly servicing nearly all in the region.[4]

Those few *independent agents* who are not patrons of the agency take it upon themselves to defend their rights. The last step in Nozick's origin story of the state comes when the agency prohibits independents from exacting justice through unilateral use of force[5]. Independents, in the course of enforcing their own rights, impose unnecessary risks on patrons of the agency, meaning the private enforcement of justice constitutes a 'public wrong' (Osterfeld 1983). For instance, a well-meaning independent may severely over-estimate the level of punishment another deserves and kill a client of the agency. Yet in prohibiting independents

[3] Nozick claims private protective agencies will attend to 'all functions of detection, apprehension, judicial determination of guilt, punishment, and exaction of compensation' (13).

[4] Nozick does not refer to this agency as a *minimal* state since not all have purchased protection from the monopoly.

[5] This signifies the move from the ultra-minimal to the minimal state. Nozick claims a 'necessary condition for the existence of a state is that it announce that, to the best of its ability [...] it will punish everyone whom it discovers to have used force without its express permission' (24).

from meting out punishment, the agency effectively violates the rights of independents. As a result, compensation is in order. In particular, Nozick contends the agency must at a minimum provide free protection to independents. Additional compensation may be required, although this is a matter Nozick is notoriously slippery on.[6]

At this point, the agency now has many of the familiar features of the state. The dominant protection agency has a monopoly on the use of force and provides protection to all residing in a particular location. Payments are made to the state but this does not violate self-ownership rights (or so Nozick would argue) since original members of the agency paid voluntarily. Thus a *minimal* state has emerged from a series of voluntary exchanges between morally motivated individuals and protection agencies in the Lockean SON. This origin story stands as a possibility proof, intended for the individual anarchist. Exploitation and oppression are not, as the anarchist contends, intrinsic features of the political realm. The state can come about through a series of voluntary transactions in which rights are respected.

Others have taken Nozick to be up to something substantially grander.[7] Gaus (2011a, b) takes Nozick to provide a *fundamental explanation* of the realm of the political. Nozick's invisible-hand story shows how statehood (or some version of *minimal-statehood*) can be explained without an appeal to concepts such as political authority or sovereignty. Gaus further contends that if Nozick's origin story is successful, it shows how political legitimacy is an emergent property of ultra-minimal states generally, and moreover, the legitimacy of the ultra-minimal state arises from "non-political interactions, and emerges upon a wide variety of social states" (p. 14). In a similar fashion, Bader (2017) develops an account of 'counterfactual justification' inspired by both Nozick's SON story as well as Nozick's entitlement theory. On Bader's account, Nozick's justificatory strategy is to appeal to what "happens in the closest possible world in which circumstances are ideal," making it necessary that the account both use as its starting point the moralized Lockean SON and take the "form of a hypothetical explanation based on an invisible- hand mechanism" (p. 11). While our primary goal in this paper is not to provide an account of legitimacy nor defend the justificatory strategy taken by Nozick, we briefly return to these claims in Sect. 5, after first developing a game-theoretic representation of the Lockean SON.

3 Modeling the march to the minimal state

In this section we develop a game-theoretic representation of the Lockean SON as well as Nozick's invisible-hand story. First, however, a few words about our game-theoretic approach. Game theory is a branch of mathematics used to understand strategic behavior in a variety of contexts. A *game* consists of *players* who can

[6] See, for instance, Hyams (2004) and Wood (1978).

[7] Nozick himself appears to believe his origin story accomplishes multiple philosophical tasks, noting it provides a fundamental explanation of the political as well as a refutation of the individual anarchist.

employ a variety of *strategies*. The combination of strategies selected by players in a game determines the payoff for those involved. Game theorists then focus on how rational individuals act in such settings. Since much theorizing about the SON considers how individuals learn to behave in various strategic scenarios, game theory is especially illuminating. As a result, game theory has been employed by many contemporary social contract thinkers such as Kavka (1986, manuscript) and Hampton (1988).[8] For our purposes, game theory allows us to understand when certain social arrangements are, or are not, stable. Since Nozick's origin story implicitly assumes certain social arrangements are stable (the minimal state), while others are not (the Lockean SON), game-theoretic tools are particularly apt.

3.1 Locke's state of nature

We begin in the Lockean SON where there is disagreement among persons about what justice requires. In Nozick's description of the Lockean SON, the vast majority of individuals comply with the LN but are nonetheless biased. This leads them to systematically interpret rights claims in their favor and overestimate the compensation they are owed. Disagreement between well-meaning, moral individuals is expected, and violent conflict, as a result, can easily ensue.

Yet to claim that the result of disagreement is inevitably a long cycle of retaliations is too quick. While all individuals are self-biased, this does not entail *all* will be equally eager to aggressively impose their judgments on others. Some may advocate for a level of punishment or compensation but, in the face of disagreement, opt for a compromise. Others, less so. These latter individuals will stubbornly punish and seek compensation from their counterpart in cases of disagreement. If their counterpart is similarly tenacious, retaliation does appear inevitable, as the punished individual will now seek redress for what they perceive to be an unjustified overreaction. Both sides will feel wronged, fueling animosity and future conflict.

We contend this situation is best represented as a *conflictual coordination game*, similar to the so-called hawk-dove game.[9] When individuals are self-biased the result is disagreement: one individual thinks they are owed some amount in compensation, while another protests that this figure is an overestimate.[10] If both are similarly disposed to resolve the disagreement peacefully (both are Concessive), it is reasonable to assume some sort of compromise will be reached (this is relaxed in Sect. 3.3). If both are stubborn (both are Steadfast), conflict breaks out and Nozick's cycle of retaliations is set in motion. When an aggressive individual interacts with one inclined to resolve peacefully, the aggressive individual gets their way. Thus, the best outcome for an individual is for them to remain steadfast while their counterpart is concessive, and the worst outcome is for both parties to dig their heels

[8] For more recent work, see Skyrms (1996, 2004), Binmore (2005), Bruner (2015, 2018) and Chung (2016).

[9] Other games, such as a bargaining game, could be appealed to as well. We for simplicity focus primarily on the hawk-dove game.

[10] We do not focus on punishment. The addition of some moderate level of punishment to Tables 1 and 2 does not significantly change qualitative results.

in. The second-best outcome is for both parties to make a concession, and the third-best outcome involves the focal individual conceding to the demands of their more aggressive counterpart. This is summarized in Table 1.

Note individuals do best to engage in behavior that complements the behavior of their counterpart. If one's counterpart plays Concessive, then the best course of action is to play Steadfast. Likewise, conceding is in one's best interest when paired with an aggressive individual. Thus, (Steadfast, Concessive) and (Concessive, Steadfast) are both equilibria in this game.

We convert Table 1, featuring ordinal utilities, into a payoff table involving cardinal utilities (Table 2). This requires a few additional assumptions. First, we assume individuals are equally likely to be either victim or perpetrator. Recall that the strategic scenario of interest is one in which a rights violation has occurred and the victim must take it upon themselves to either punish or seek compensation from the perpetrator. Furthermore, since individuals in the Lockean SON are taken to be well-meaning individuals attempting to adhere to the LN, we assume rights violations are due to accidents or reasonable disagreement regarding how to interpret the LN (this is relaxed in Sect. 4). As a result, individuals are equally likely to be victim or perpetrator. If victim, the agent demands some level of compensation, C_V, from the perpetrator. The perpetrator, on the other hand, believes they owe the victim a level of compensation, C_P, where $C_V > C_P$. Thus, the difference between these two figures $(C_V - C_P)$ reflects how 'self-biased' individuals are. We refer to this level of self-bias as b. Now consider the interaction between an individual playing Steadfast and another playing Concessive. Half of the time the individual playing Steadfast will be in the role of victim. She will demand to be compensated at the level of C_V and, since her counterpart plays Concessive, will succeed in acquiring this payment. Likewise, when placed in the role of perpetrator, the individual will successfully pay the victim the lesser amount of C_P. Thus, on average an individual playing Steadfast will secure a payoff of $.5(C_V - C_P)$ from their passive counterpart. Note that this is simply one half of b $(.5(C_V - C_P) = 0.5b)$. Likewise, the payoff associated with the agent playing Concessive in this case is $-0.5b$.

Table 1 Conflictual coordination game (ordinal utilities)

	Steadfast	Concessive
Steadfast	I, I	IV, II
Concessive	II, IV	III, III

Table 2 Conflictual coordination game (cardinal utilities)

	Steadfast	Concessive
Steadfast	$-c, -c$	$0.5b, -0.5b$
Concessive	$-0.5b, 0.5b$	0, 0

In a similar vein, when two individuals both playing Concessive meet they come to a compromise, which, on average, privileges neither party's claims. In other words, the victim receives a payment of $(C_V + C_P)/2$ from the perpetrator. On average, then, agents playing Concessive receive a payoff of 0 when interacting with fellow accommodating individuals. Finally, when two agents playing Steadfast interact, conflict breaks out. Keeping with Nozick's description, conflict is devastating, leaving both with a payoff of $-c$.

Once again, we see that our conflictual coordination game allows for two pure-strategy equilibria in which players perform complementary acts. However, a mixed equilibrium involving both individuals utilizing the available acts with some non-zero probability is also possible. In particular, a mixed equilibrium exists where both agents play Steadfast with probability $b/2c$ and Concessive with probability $\frac{2c-b}{2c}$. Note that this equilibrium is not efficient, as all would do better if both play Concessive (at the mixed equilibrium both agents secure a payoff of $-b^2/4c$ which is less than 0).

How will a group or whole community of individuals in the SON behave when confronted by the game in Table 2? As we demonstrate, a stable mixture of strategic behavior is expected. To see this, consider the case in which all individuals play Concessive. In this case, whenever two members of the community fall into disagreement, they both opt for a compromise, and violence is avoided. As a result, all secure an average payoff of 0. Now consider the introduction of a lone individual who instead plays Steadfast. When paired with any of the native members of the population, this invader will always get her way, and thus will claim an average payoff of $0.5b$. Individuals thus have incentive to change their strategy from Concessive to Steadfast. Similarly, consider behavior at the other extreme. When all play Steadfast, violence is a guarantee, and thus an individual does best to switch from Steadfast to Concessive.

While both extremes are untenable, there is a mixture of individuals playing Steadfast and Concessive that is in fact stable.[11] Call this the *SON equilibrium*. This equilibrium corresponds to the mixed strategy equilibrium discussed above. In other words, the proportion of individuals playing Steadfast and Concessive at the SON equilibrium is equal to the probability an individual at the mixed strategy equilibrium plays Steadfast and Concessive, respectively.[12] We follow the game-theoretic literature, and refer to such a stable mixture of behavior as a *polymorphic* equilibrium.[13] At this polymorphic equilibrium, the proportion of individuals utilizing Steadfast is $b/2c$. If the proportion of Steadfast types for some reason increases above this threshold, those playing Concessive will now outperform those utilizing Steadfast, resulting in an uptake of the Concessive strategy, thereby

[11] To clarify, we draw on stability concepts from evolutionary game theory. The arrangement in which all play Steadfast is not stable because Steadfast is not an evolutionarily stable strategy (nor is Concessive). See Sandholm (2010) for an overview of stability concepts from evolutionary game theory.

[12] In the terminology of evolutionary game theory, this stable mixture of individuals playing Steadfast and Concessive is an evolutionarily stable state.

[13] To be precise, a polymorphic equilibrium consists of a stable mixture of various strategic types in the population.

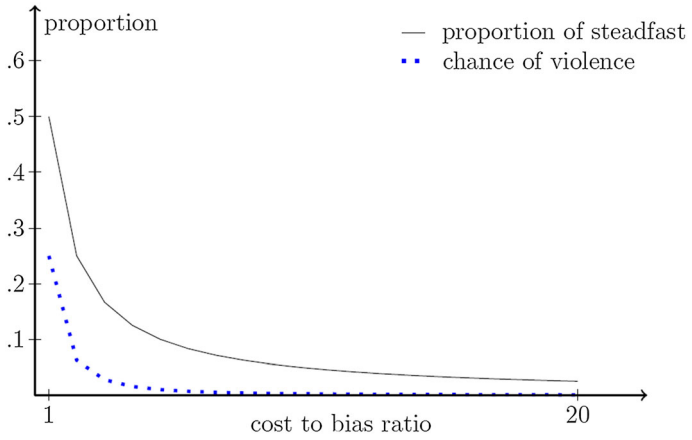


Fig. 1 Proportion of Steadfast and chance of violent conflict at the SON equilibrium for various c to b ratios

pushing the population back to the polymorphic equilibrium.[14] In the Lockean SON, play converges on a stable mixture of those playing Steadfast and Concessive.

What determines the proportion of Steadfast and Concessive types at the SON equilibrium? Note that as the cost of conflict (c) increases, the proportion of those playing Steadfast ($b/2c$) decreases; as the level of bias (b) decreases, the proportion of those playing Steadfast at equilibrium goes down. We should expect conflict to be at a minimum when individuals are not particularly biased and the cost of conflict is rather large: conditions closely corresponding to Nozick's informal characterization of the SON. Recall Nozick assumes that when disagreements turn violent the result is often a cycle of retaliations. Furthermore, while Nozick doesn't explicitly speak to just how biased individuals will be in the SON, the fact that individuals are assumed to be well-meaning and compliant with the LN suggests that individuals shouldn't be excessively biased, even when adjudicating their own affairs. Taken together, we have good reason then to think that in Nozick's imagining of the SON, the proportion of Steadfast types in the population will be particularly small, meaning violent conflict is by no means the norm. For a better sense of the composition of types at the SON equilibrium, see Fig. 1. Note that when the cost of conflict (c) is ten times the level of bias (b), 5% of the population plays Steadfast, meaning on average less than one quarter of a percent of pairings result in violent conflict.

In close, since violence is somewhat uncommon individuals may not have incentive to patronize protection agencies. We now turn to this question of whether individuals benefit from the introduction of private protection firms.

[14] Individuals can come to change their behavior via individual or social forms of learning.

3.2 Firms and free-riders

Nozick provides an invisible-hand story of agency competition which eventually results in a dominant protection agency. Many philosophers and social scientists have criticized this account of how competition between agencies will play out. Some, for instance, have contended the private provision of protection will be prohibitively costly as violent conflicts between agencies will drive up prices.

In what follows, we consider the best case scenario in which there is just a single agency and this agency prices their services at the lowest possible rate. We then explore whether individuals would have incentive to patronize this lone agency. Complicating factors such as agency competition and the creation of a 'confederacy of agencies' are not addressed in this paper. Instead, we explore whether individuals would have incentive to patronize a protection agency in the *ideal* case where a lone agency offers its services at the lowest possible price. If the minimal state is a far cry in these extremely favorable circumstances, we contend that the minimal state is very unlikely to arise in conditions more closely approximating those assumed by Nozick.

We modify the conflictual coordination game from the previous section by introducing an additional strategy: purchase services from the protection agency (Agency). While this comes at some cost, the benefits are rather obvious. When interacting with fellow clients, an individual will be able to rely on the agency to adjudicate matters and thus avoid the concessions and conflict characteristic of the SON. Furthermore, the agency will likewise adjudicate disputes when a client interacts with an independent who has not purchased services from the agency. In this case, the agency will take it upon themselves to protect clients from the behavior of independents, ensuring they won't be attacked in the future if independents are not satisfied with how a particular dispute was resolved.

As mentioned, we consider the best case scenario involving just one agency pricing their services at the lowest possible level. Will, in such favorable circumstances, agents in the Lockean SON select to patronize the agency? To answer this question we must specify payoffs. Consider the interaction which takes place between two members of the agency: when disagreement arises, the agency resolves the dispute and forces one member to compensate the other by some amount. Since both individuals have the same chance of being victim or perpetrator, half of the time an individual doles out funds to their counterpart, and half of the time they receive said funds from their counterpart. Thus, when two clients interact,

Table 3 Conflictual coordination game with protection agency

	Steadfast	Concessive	Agency
Steadfast	$-c, -c$	$0.5b, -0.5b$	$0, -e$
Concessive	$-0.5b, 0.5b$	$0, 0$	$0, -e$
Agency	$-e, 0$	$-e, 0$	$-e, -e$

they on average receive a payoff of zero. Finally, we must take into account the fact that agents patronizing the agency pay some membership fee. Let us assume that on average individuals get into N disagreements in each payment period. Given this fact, the agency then prices their services for a given payment period at the lowest possible price which is still economically viable. Let this minimally expensive membership fee be eN. Thus, the client on average pays a small cost of e for each disagreement they are dragged into. We modify Table 3 to reflect this.

What occurs when clients and independents meet? To answer this question, first consider the related situation in which clients of *two different agencies* fall into disagreement. In this case, Nozick contends that if the two agencies have different procedures for adjudicating disputes, the agencies either fight or come to some kind of compromise as to which of the two procedures will be utilized to settle the original disagreement between their clients. Likewise, when an independent falls into disagreement with a client of the agency, the agency and independent may also have different procedures for resolving disputes. In this case, however, the agency (with its superior power and artillery) can easily impose their procedure on the lone individual. Thus, whenever an independent and client of the agency butt-heads, the conflict is resolved according to the agency's conflict-resolution procedure.[15]

Yet what prevents the agency from exploiting a lone independent, stripping them of their possessions under the guise of resolving a dispute? Nozick addresses this concern and explicitly rules out such 'predatory agencies' since their behavior would constitute an obvious violation of the LN[16]. A related worry is that biased agencies would systematically favor their clients. In this case, when clients and independents disagree as to how much is owed in compensation, the agency always advocates for and enforces the level of compensation which suits their client regardless of the facts. Nozick suggests such a protection agency runs the risk of undercutting its legitimacy if it becomes a state. This is due to the fact that individuals would come to patronize the agency not because they are attracted to its services, but to instead avoid mistreatment. As a result, clients may come to view themselves as *victims* of the agency and fail to ever conceive of themselves as true citizens of the minimal state (Nozick 1974, p. 17). Nozick further claims this is doubly problematic: a successful state requires a certain level of voluntary cooperation and compliance. If a large portion of the community view the state as illegitimate, cooperation is doubtful and the state will fail to carry out many of its central goals, or so Nozick claims. Thus, the inclusion of predatory and biased agencies in the SON is a problem for Nozick's libertarian project: such agencies would succeed at attracting customers, but for all the wrong reasons (we return to this in 3.3 and explore more carefully whether these biased agencies will, in fact, successfully attract customers. As it turns out, such agencies can recruit large swaths of the population, but only under certain conditions).

[15] Note the agency is not prohibiting independents from enforcing their rights as independents can use force on other independents. However, if an independent's use of force against a client is not in-line with the agency's procedure the agency rectifies the situation.

[16] In other words, both individuals and agencies make a 'good faith [effort to] act within the limits of Locke's [LN]' (Nozick, p. 17).

We now return to the game of Table 3. When an independent and a client of the agency interact, the agency does not behave in a biased fashion and thus the agency's way of resolving the dispute on average favors neither party. Independents receive a payoff of zero when interacting with clients. Clients likewise receive a payoff of zero in addition to their perfunctory membership fee of $-e$. Finally, when two independents meet, the agency is not involved and thus payoffs mirror those outlined in Table 2. This completes our description of the game. We now discuss how play will unfold.

First, we consider the case in which no one patronizes the firm. A stable mixture of Steadfast and Cocnessive individuals is possible. In this case, a proportion of the population ($b/2c$) play Steadfast, while the remaining individuals play Concessive. Note that this mixture is identical to the mixture of individuals playing Steadfast and Concessive at the SON equilibrium from the previous section. Individuals in this case secure an average payoff of $-b^2/4c$. The payoff associated with playing Agency ($-e$) is less than this average payoff when $b^2/4c < e$. In other words, when the cost of membership is greater than $b^2/4c$, an individual playing Agency cannot invade the polymorphic equilibrium consisting only of Concessive and Steadfast individuals.

Alternatively, there exists a polymorphic equilibrium consisting of all three strategies. The proportion of individuals playing Steadfast is $2e/b$, while the proportion of those patronizing the firm is $1 - 4ec/b^2$. Note that for these proportions to be in the unit interval, the following conditions must be met: $e < b/2$ and $e < b^2/4c$.[17] Recall from above that the SON equilibrium exists when $e > b^2/4c$, meaning these two social arrangements are never both stable for the same set of parameter values.[18]

Yet while some may purchase protection from the agency at the polymorphic equilibrium, this is not to say that *most* will join the protection agency. For one, if all subscribe to the protection agency individuals have incentive to unilaterally change their strategy and become independent. Independents do better than those playing Agency as they reap the benefits of the agency's conflict resolution services without having to pay a membership fee. Put another way, independents *free-ride* on the contributions of clients[19]. When dealing with clients, independents avoid both

[17] Note that Agency is not strictly dominated by Concessive when $e < b/2$. Furthermore, $b/2$ is greater than $b^2/4c$ whenever $c > b/2$, which is assumed to hold if the Lockean SON is to be a conflictual coordination game. Thus e must be less than $b^2/4c$ for the polymorphic equilibrium to exist.

[18] For completeness, we also consider the polymorphic equilibrium consisting of Agency and Concessive as well as the polymorphic equilibrium consisting of Agency and Steadfast. The former arrangement is not stable as those playing Agency will always do worse than those playing Concessive ($-e < 0$). As for the latter arrangement, it is easy to show that the proportion of those playing Steadfast is equal to e/c. A Concessive invader will secure a payoff of $-eb/2c$, which is greater than the average payoff of the natives ($-e$) when $c > b/2$. Note that this inequality is a necessary condition for the underlying Lockean SON to be a conflictual coordination game. Thus the polymorphic equilibrium does not exist for parameter values of interest.

[19] Nozick does not appear to be aware of this free-rider problem, although he identifies a *different* free-rider problem in *ASU* in which clients purposefully disassociates themselves from the agency in order to be 'bought off'.

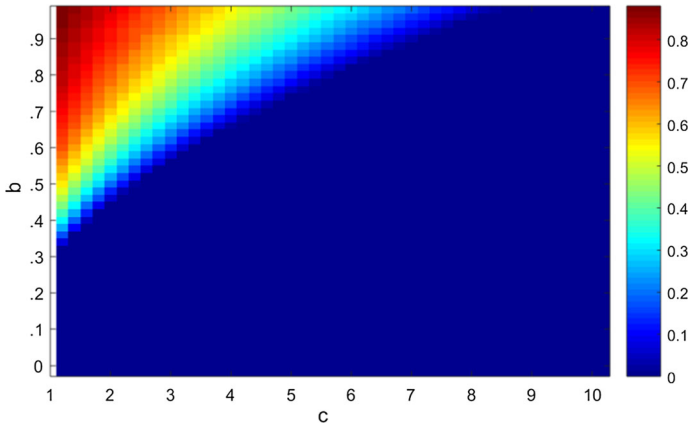


Fig. 2 Proportion of individuals purchasing protective services from the agency ($e = 0.03$)

the violent conflict and excessive concessions that are characteristic of the Lockean SON.[20]

This unnoticed free-rider problem prevents individuals in the SON from all subscribing to the protection agency.[21] Further, a significant fraction, if not a majority, of individuals do not subscribe to the protection agency at the polymorphic equilibrium. Recall the proportion of individuals patronizing the agency is $1 - 4ec/b^2$. For relatively high values of c and modest to low values of b, a small proportion of individuals play Agency at this polymorphic equilibrium (Fig. 2).

To further illustrate how unpopular the protection agency will be, we consider the minimum 'cost-of-violence to bias' ratio (c/b) necessary to ensure that 50% of the population patronizes the agency. If, for instance, this ratio is 5 for a given set of parameters, then when the cost of conflict is five times that of bias, 50% of the population patronizes the agency at the polymorphic equilibrium. Ceteris paribus, if the cost of violence increases, fewer than half of the population will play Agency at the polymorphic equilibrium. As Fig. 3 indicates, this ratio need not be particularly high, further suggesting that few individuals will come to patronize the agency when placed in Nozick's imagining of the SON.

Taken together, these results suggest Nozick's origin story is stalled: protection agencies may exist, but often only attract a small fraction of the community (if any at all). Could the agency 'buy off' independents? This is infeasible in most cases since (1) the proportion of independents is high (often exceeding 50%) and (2) the cost of membership is by assumption as low as economically feasible. Thus the agency would not be able to easily extend its services to many independents without

[20] One may argue that this free-rider problem is avoidable if agencies extract a payment from independents. While clients have a contract with the agency, the unaffiliated have no such agreement. Thus payment extraction would, to Nozick, amount to theft.

[21] Peter Vanderschraaf has communicated via email to me that he has independently discovered this free-rider problem.

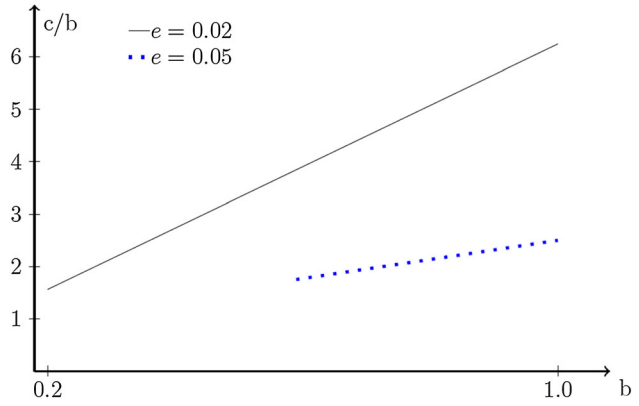


Fig. 3 Minimum c to b ratio (y-axis) such that no more than 50% of community at polymorphic equilibrium plays agency. Linear equation is $c/b = 6.25b$ and $c/b = 2.5b$ for the blue and orange line, respectively

compromising quality. Furthermore, any attempt to raise prices on clients would shift the proportion of individuals patronizing the agency at equilibrium, resulting in even fewer clients.

3.3 Robustness check

We've argued that while protection agencies may thrive, their success is capped by a rather low ceiling. We now consider a few natural modifications of our baseline model and contend none of these alter our central result.

First, we consider what occurs if clients can select to handle conflicts with independents on their own by playing either Concessive or Steadfast against independents. By resolving disputes themselves, independents are no longer able to 'free ride' on clients of the agency. We consider the best case scenario in which protection agencies provide clients a refund for each case of disagreement clients themselves resolve.

Recall in Sect. 3.2 that individuals on average paid e for each interaction the agency helped resolve. We now consider the case in which clients of the agency retain e each time they take it upon themselves to resolve disputes with independents. We can somewhat straightforwardly prove that at the polymorphic equilibrium, those who patronize the agency will never have incentive to attempt to resolve disputes by themselves when their counterpart is unaligned. To see this, consider the following. At the polymorphic equilibrium in which all three types are present (Steadfast, Concessive and Agency), everyone receives the same expected payoff of $-e$. Independents receive a payoff of 0 when interacting with those aligned with the agency, which means the expected payoff of interacting with those not aligned with the agency must be less than $-e$ in order to ensure that their overall expected payoff is $-e$. Hence, an individual aligned with the agency who instead decides to play either Concessive or Steadfast against an individual not aligned with

Table 4 Conflictual coordination game with Biased Agency

	Steadfast	Concessive	Agency
Steadfast	$-c, -c$	$0.5b, -0.5b$	$-$ b/2, $b/2 - e$
Concessive	$-0.5b, 0.5b$	0, 0	$-$ b/2, $b/2 - e$
Agency	$b/2 - e, -b/2$	$b/2 - e, b/2$	$-e, -e$

the agency will receive a payoff of less than $-e$. Thus, clients of the agency will never have incentive to resolve conflicts with independents on their own.

We now consider a different modification to our baseline story. We assume that when two Concessives meet they incur a cost in the course of attempting to resolve their disagreement. This is not to say that violence erupts, just that it may take some time and effort to amicably arrive at a settlement. This may in turn make the services provided by the protection agency appear all the more attractive. We modify the game in Table 3 to reflect this new cost. The inclusion of this cost does tend to tip the scales in favor of the protection agency, although for the relevant parameters (high c, low b), it is still the case that sizable swaths of the population do not patronize the agency. For instance, when $c = 2$, individuals are moderately biased ($b = 0.1$) and the transaction cost, f, is equal to the cost of firm membership ($e = 0.02$), around 5% of individuals will select to patronize the agency (as opposed to 0% when f is set to 0). When f is increased to twice the cost of the membership fee, the proportion of those patronizing the agency jumps to 50%. When f is increased to four times the value of e, this number becomes 74%. Thus, even when the cost associated with amicably resolving a dispute is rather high, a significant proportion of the population still chooses to remain independent.

Finally, we explore the 'biased' agencies briefly discussed in Sect. 3.2. Recall that agencies may systematically favor their clients when resolving disputes. While we argued the inclusion of such agencies in Nozick's origin story would undermine Nozick's philosophical conclusion, we nonetheless consider how such agencies will fare. Briefly, we find that biased agencies can sometimes recruit large swaths of the community. However, agencies can still routinely fail to attract customers, reinforcing our above main finding.

Biased agencies systematically favor their clients. This means that when their client has a disagreement with a non-client and is in the role of 'victim' the agency ensures their client receives C_V (where $C_V > C_P$). Likewise, when their client is in the role of 'perpetrator' the agency forces their client to pay the smaller amount of C_P to the victim. This means the average payoff a client of the agency receives when interacting with a non-client is $.5C_V - .5C_P = .5b$. Likewise, we calculate the average payoff for the non-client to be $-.5b$ (see Table 4).

Note that when $e < b/2$, the scenario where all subscribe to the protective agency is stable. When $e > b/2$ this arrangement is no longer stable as individuals have incentive to play Steadfast or Concessive. In this case there are two possible stable arrangements. The first corresponds to the SON equilibrium from Sect. 3.1. Namely, at equilibrium a portion of the community plays Steadfast ($b/2c$), while the

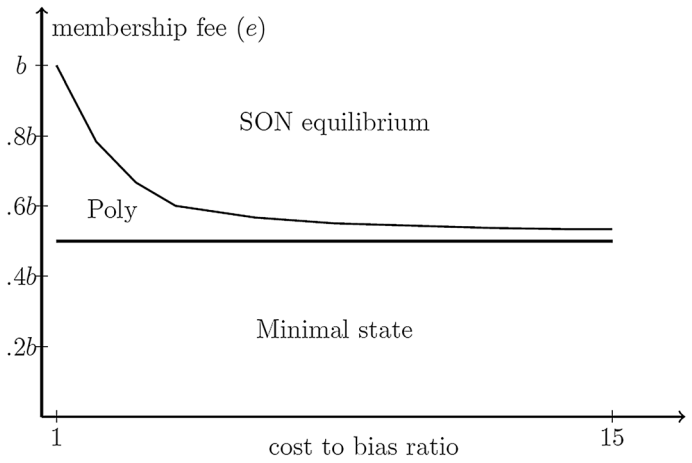


Fig. 4 Regions of parameter space that support the SON equilibrium, the Polymorphic equilibrium involving all three strategies, and the Minimal state (the monomorphic population where all purchase protection from the agency). c to b ratio displayed on the x-axis and the cost of protection (e) on the y-axis as a proportion of b. Figure is for $b = 1$

remaining individuals play Concessive. This is stable when $e > b^2/4c + b/2$.[22] The second stable arrangement is a mixed equilibrium involving all three strategies. In this case, individuals play agency with probability $1 - 4ce/b^2 + 2c/b$ and Steadfast with probability $2e/b - 1$. A necessary condition for this polymorphic equilibrium is $b/2 < e < b^2/4c + b/2$, meaning this equilibrium and the SON equilibrium will never both be stable for the same set of parameter values.

Additionally, for parameter values of interest, none of the other polymorphic equilibria are stable.[23] This means play results in one of three possibilities: the SON equilibrium, the polymorphic equilibrium involving all three strategies and the equilibrium where all play Agency. Finally, only one of these three social arrangements is stable for a given set of parameter values. Figure 4 illustrates this graphically.

For the transition to the minimal state to occur, it appears agencies must exhibit a significant bias in favor of their clients. Yet Nozick excludes Biased Agencies from his origin story. This can be clearly seen, for instance, in his discussion of Agency-

[22] Note that the average payoff at this mixed strategy equilibrium is $-b^2/4c$. For this mixed strategy equilibrium to be stable, individuals must not have incentive to patronize the agency. This occurs when $b/2 - e$ is less than $-b^2/4c$, which holds when $e > b^2/4c + b/2$.

[23] Consider the polymorphic equilibrium consisting of Agency and Concessive. In this case, Agency strictly dominates Concessive when $e < b/2$ and Concessive strictly dominates Agency when $e > b/2$. Furthermore, at the polymorphic equilibrium involving Agency and Steadfast, the proportion of individuals playing Steadfast is $(2e - b)/2c$, which is in the unit interval when $b/2 < e < b/2 + c$. The average payoff at this polymorphic equilibrium is $be/2c - b^2/4c - e$, while an invading agent playing Concessive receives a payoff of $-b/2$. Those playing Concessive secure a lesser payoff when $e < b/2$ holds. Note, however, that e must be greater than $b/2$ for the proportion playing Steadfast to be in the unit interval. Thus this polymorphic equilibrium is not stable.

Agency conflict. Here Nozick assumes agencies are impartial and attempt to rectify potentially dangerous situations by appeal to their preferred conflict resolution procedure. Conflict between agencies, then, is not the result of bias but instead due to the use of different conflict resolution procedures. Furthermore, when discussing the de facto monopoly, Nozick claims that it will be 'generally known' that the procedures utilized by this Agency are both reliable and fair. Thus while the results of this section indicate the free-rider problem can be avoided, this is only possible if we deviate from some of the assumptions Nozick makes about these Agencies.

4 The problem of coordination

As we have seen, one of the central problems confronting individuals is that of *coordination*. Individuals have incentive to coordinate on complementary behaviors in the SON. Unfortunately, not all interactions at the SON equilibrium involve participants taking complementary acts, and thus conflicts arise, albeit infrequently. Can individuals learn to better coordinate their behavior and completely avoid conflict? The answer is a resounding 'yes', but to see this we must first introduce a new kind of strategy.

4.1 Correlated conventions

Consider the following situation. When you and another have a disagreement, a coin is flipped. If you call 'heads' and the coin lands on heads, you remain steadfast while your counterpart concedes. If the coin lands on 'tails', the roles are reversed: you now concede while your counterpart receives what they have claimed. Given this description, neither have incentive to unilaterally deviate and behave differently in response to the result of the coin-flip. If the coin comes up 'heads' and your counterpart instead refuses to concede, conflict breaks out. She would have done better to 'stick with the plan'. This is an instance of a *correlated equilibrium* (Aumann 1974).

Correlated equilibria typically involve a public signal that both individuals condition their behavior on (such as a coin flip). Skyrms (1996) has shown those employing conditional strategies such as 'Concede if heads, remain Steadfast if tails' can overtake a community currently at an inefficient polymorphic equilibrium (such as the SON equilibrium from Sect. 3.1).[24]

Yet individuals need not appeal to some external randomization device. Coordination can be facilitated by attending to certain asymmetries naturally present in the strategic scenario of interest[25]. For instance, individuals in the SON inhabit one of two roles: perpetrator or victim. Individuals then condition their behavior on their present role (which, as we have been assuming, is randomly

[24] See also Gaus (2011b) as well as Vanderschraaf (1995, 2001, 2018).

[25] See Maynard-Smith's (1982) work on the so-called 'property equilibrium'.

determined). This allows for the following conditional strategy: 'Concede if Perpetrator and remain Steadfast if Victim.' Note that just as in the coin-flip case, individuals can perfectly coordinate their behavior by conditioning on seemingly random factors. Two individuals playing the above strategy will never both hold steadfast.[26] Thus, the situation in which both individuals play the above conditional strategy is an equilibrium. Likewise, the arrangement where the perpetrator always gets their way when resolving a dispute and the victim concedes is also stable. Individuals using these conditional strategies can infiltrate a community currently at the SON equilibrium from Sect. 3.1. These invaders do just as well when playing against natives as natives do against themselves. When two invaders meet, however, they are able to perfectly coordinate, thereby securing a higher payoff than the natives.

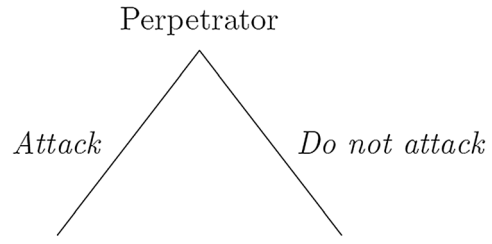
If individuals can utilize these conditional strategies, conflict in the SON can potentially be completely avoided. Furthermore, if individuals are able to peacefully resolve their disagreements, then protection agencies may play a limited role in the SON. In fact, at the correlated equilibrium in which perpetrators concede to the demands of the victim, individuals will not have incentive to patronize a protection agency. Individuals using the conditional strategy on average receive a payoff of 0 when interacting with a counterpart using the same conditional strategy. This is greater than the payoff those patronizing the agency secure $(-e)$. Thus, at the correlated equilibrium, those patronizing the agency will not be able to gain a foothold in the community. Further note that at the polymorphic equilibrium from Sect. 3.2 in which all three strategies (Steadfast, Concessive, Agency) are present, conditional strategies can thrive, as once again these invaders do as well against natives as the natives do against themselves $(-e)$, but attain a higher payoff when interacting with those using the same conditional strategy (0). Thus not only is the correlated equilibrium impervious to the introduction of protection agencies, conditional strategies can infiltrate the mixed equilibrium from the previous section, pushing out the agency entirely.

4.2 Leaving Eden

Our rosy picture of the Lockean SON is predicated on a strong assumption: individuals attempt to abide by the LN. If relaxed, individuals violate the rights of others when it is to their benefit. Will a state of malice and enmity be unavoidable? We introduce a slightly modified version of the game in Table 1 to address this (see Fig. 5). When two individuals interact, one can select to 'attack' the other, securing some of their counterpart's property or power.[27] If such a rights violation occurs, the second individual can attempt to punish or seek compensation from the perpetrator. Once again, disagreement arises as the victim overestimates and the perpetrator

[26] Of course, the situation is complicated if it is unclear who is victim and perpetrator.

[27] We follow Vanderschraaf (2006) and assume attacks secure either goods or physical powers. We assume only one individual can currently take advantage of the other, which is determined by chance.



	S	*C*
Steadfast	I, I	V, II
Concessive	II, IV	III, III

Fig. 5 Extensive form of SON game (ordinal payoffs). Perpetrator is row player and Victim is column player

underestimates the harm suffered by the victim.[28] We once again assume this strategic scenario is best captured by a conflictual coordination game such as hawk-dove. Furthermore, we assume that the best outcome for the victim is for the perpetrator to not attack the victim, while the best outcome for the perpetrator is to levy an attack and remain steadfast while the victim concedes. See Fig. 5 for the complete ranking of outcomes.

The set of available strategies is slightly more complicated than what we have seen thus far. First, individuals must decide whether to initiate a rights violation or not when in the role of perpetrator. Second, individuals must determine whether to play Steadfast or Concessive as perpetrator and victim. 'Attack and remain steadfast if Perpetrator, concede if Victim' is a possible strategy. If used by all, potential perpetrators will not have incentive to concede given the passive behavior of their counterpart. Those in the role of perpetrator will select to attack as they will not be punished by their counterpart. Thus, all individuals select to violate the rights of others when it is to their benefit and individuals do not bother to attempt to seek redress.

Yet this is not the only stable outcome of the above game. Consider the strategy 'Do not attack and concede if Perpetrator, remain steadfast if Victim'. In this case, potential perpetrators do not have incentive to violate the rights of others since such violations will be meet with force.[29] Namely, the victim will aggressively demand compensation and punish those who violate her rights. Note that this is not based on an incredible threat on the part of the victim. The victim has incentive to pursue compensation to the fullest extent if she believes her counterpart will easily concede.

[28] We expect the perpetrator *severely* underestimates damage.

[29] In particular, this is a subgame perfect equilibrium as play constitutes a Nash equilibrium in each subgame.

These two possible endpoints of our game correspond to two informal (and seemingly conflicting) descriptions Locke provides of the SON.[30] Locke at times characterizes the SON as a place of tranquility while at other points as one rife with ill-will and enmity. While some scholars accuse Locke of blatantly contradicting himself, others contend Locke was providing an account of *possible* states of nature. John Simmons (1989, 1993), for instance, views these conflicting descriptions as characterizations of the best and worst life could be like in the SON.[31] When individuals tend to obey the LN 'the state of nature will be one of peace and goodwill and the like; where persons disregard the law, the state of nature will be a state of enmity, malice and so on' (Simmons 1989, 459). We have shown that both of these scenarios are in fact stable: when most respect the LN, a correlated equilibrium involving no conflict is possible. When individuals are not moved by the LN, theft and rights violations are commonplace. Somewhat surprisingly, though, we have also shown how a peaceful and tranquil SON is possible even when individuals are not intrinsically motivated to follow the LN. In this case, conditional strategies ensure potential aggressors do best to not infringe on the rights of others so as to avoid punishment.

5 Conclusion

We have argued that Locke's SON is best conceived of as a conflictual coordination game where both parties attempt to settle a dispute regarding the demands of justice. On this interpretation, conflict is by no means the norm. In fact, violence occurs rather infrequently, suggesting the social characterization of the Lockean SON is a far cry from a war of all against all. Furthermore, we found that when a preponderant majority of individuals adhere to the LN, only a fraction of the community patronizes the protection agency due to a hitherto unnoticed free-rider problem. The transition to the ultra-minimal state is blocked: a large proportion—often a majority—of community-members do not purchase protective services at equilibrium.

What does this mean for Nozick's libertarian project and the legitimacy of the minimal state? Recall that Nozick's origin story played multiple roles, acting as a proof of possibility, a means of establishing the legitimacy of the minimal state and a 'fundamental' explanation of the political realm (Bader 2017; Gaus 2011a, b). Our central finding casts serious doubt on these claims. This is not to say, however, that no such possibility proof can be constructed, or no explanation of the political in terms of the non-political can be provided. Instead, we have simply shown the origin story supplied in *ASU* does not accomplish these tasks since the minimal state is not likely to emerge from the Lockean SON in the way Nozick anticipated.

[30] Note that for the game provided in Fig. 5 (which involves ordinal utilties), these two outcomes are the only sub-game perfect equilibrium.

[31] See Vanderschraaf (2006) for a similar game-theoretic treatment of Simmons' argument. Vanderschraaf and I depart, however, as he takes the Lockean SON to be best captred by an assurance game.

It is worth discussing how the findings of this paper compare to other game-theoretic models of the SON. Vanderschraaf (2006), for instance, conceives of the Lockean SON as a stag hunt, where individuals act in accordance with the LN if they have reason to believe their counterpart will behave likewise. This, however, fails to take into account what both Locke and Nozick took to be the central inconvenience of the SON. As we have shown, the game-theoretic representation of the strategic scenario in which individuals must stand as judge in their own case suggests violence can be avoided (Sect. 4.1), painting a much rosier picture of the SON than previously thought.[32]

Finally, whether individuals are inclined to obey the LN greatly affects the outcome of anarchy. As mentioned, this is consistent with John Simmons's reading of Locke and makes sense of the seemingly contradictory claims made in the course of *Two Treatises*. While conditional strategies can result in minimal conflict in the SON, theft and rights violations may abound if individuals are not inclined to obey the LN. In these circumstances, protection agencies would most likely thrive, perhaps laying the groundwork for the state.

Acknowledgemnets Thanks to Brian Skyrms, Simon Huttegger, Hannah Rubin, Hun Chung, Ten-Herng Lai, John Thrasher and audiences at the Social Dynamics seminar at UC Irvine, the Moral, Social and Political Theory group at ANU and the 2nd Meeting of the Politics, Philosophy and Economics Society at New Orleans. A special thanks to Peter Vanderschraaf for his encouragement and discussion on these matters over the past six years. I'm also grateful to a referee who encouraged me to provide a more thorough game-theoretic analysis of Locke and Nozick.

References

Aumann, R. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, *1*, 67–96.

Bader, R. (2017). Counterfactual justifications of the state. *Oxford Studies in Political Philosophy*, *3*, 101.

Binmore, K. (2005). *Natural justice*. Oxford: Oxford University Press.

Bruner, J. P. (2015). Diversity, tolerance and the social contract. *Politics, Philosophy and Economnics*, *14*(4), 429–448.

Bruner, J. P. (2018). Bargaining and the dynamics of divisional norms. *Synthese*. https://doi.org/10.1007/s11229-018-1729-4.

Chung, H. (2015). Hobbes's state of nature: a modern bayesian game-theoretic analysis. *Journal of the American Philosophical Association*, *1*, 485–508.

Chung, H. (2016). A game-theoretic solution to the inconsistency between Thrasymachus and Glaucon in Plato's Republic. *Ethical Perspectives*, *23*, 383–410.

[32] With respect to Nozick's story, Andrew Schotter (Schotter 1981) uses cooperative game theory to show the emergence of a dominant protection agency. His analysis rests heavily on the assumption that the LN is not respected.

Chung, H. (manuscript). Enough and as good or better? A formal comparison of right-libertarianism and Jonathan Quong's left-libertarianism.
Gaus, G. (2011a). Explanation, justification, and emergent properties: An essay on Nozickian metatheory. In *The Cambridge Companion to Nozick's Anarchy, State and Utopia*, Cambridge University Press.
Gaus, G. (2011b). The property equilibrium in a free society. *Social Philosophy and Policy*, *28*, 74–101.
Hampton, J. (1988). *Hobbes and the social contract tradition*. Cambridge: Cambridge University Press.
Hyams, K. (2004). Nozick's real argument for the minimal state. *Journal of Political Philosophy*, *12*(3), 353–364.
Kavka, G. (manuscript). Political Contractarianism.
Kavka, G. (1986). *Hobbesian moral and political theory*. Princeton: Princeton University Press.
Kogelmann, B., & Ogden, B. (2018). Enough and as good: A formal model of Lockean first appropriation. *American Journal of Political Science*, *62*, 682–694.
Locke, J. (1690/1980). *Second Treatise of Government*. Indianapolis: Hackett Publishing.
Maynard-Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.
Nozick, R. (1974). *Anarchy, State and Utopia*. New York: Basic Books.
Osterfeld, D. (1983). *Freedom, Society and the State: An investigation into the possibility of Society without government*. Lanham: University Press of America.
Sandholm, W. (2010). *Population games and evolutionary dynamics*. Cambridge: MIT Press.
Schotter, A. (1981). *The economic theory of social institutions*. Cambridge: Cambridge University Press.
Simmons, J. (1993). *On the edge of anarchy: Locke, consent, and the limits of society*. Princeton: Princeton University Press.
Simmons, J. (1989). Locke's state of nature. *Political Theory*, *17*, 449–470.
Skyrms, B. (1996). *The evolution of the social contract*. Cambridge: Cambridge University Press.
Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge: Cambridge University Press.
Vanderschraaf, P. (1995). Convention as correlated equilibrium. *Erkenntnis*, *42*, 65–87.
Vanderschraaf, P. (2001). *Learning and coordination: Inductive deliberation, equilibrium and convention*. Abingdon: Routledge.
Vanderschraaf, P. (2006). War or peace? A dynamical analysis of anarchy. *Economics & Philosophy*, *22*, 243–279.
Vanderschraaf, P. (2018). *Strategic justice: Convention and problems of balancing divergent interests*. Oxford: Oxford University Press.
Wood, D. (1978). Nozick's justification of the minimal state. *Ethics, 88*(3), 260–262.